

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 1, 2001, 15:47:01 ; Search time 210.42 Seconds
(without alignments)
10.238 Million cell updates/sec

Title: US-09-331-631a-1_COPY_186_248

Perfect score: 355
Sequence: 1 KRDPQOREYEDCRRRCEQOE.....MMNPORGSGRGVEGEEDOS 63

Scoring table:
BLOSUM62
Gapox 10.0 , Gapext 0.5

Searched: 268485 segs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: A_Geneseq_36: *
2: /SIDSL/gcgdata/geneseq/geneseq/AA1980.DAT: *
3: /SIDSL/gcgdata/geneseq/geneseq/AA1981.DAT: *
4: /SIDSL/gcgdata/geneseq/geneseq/AA1982.DAT: *
5: /SIDSL/gcgdata/geneseq/geneseq/AA1983.DAT: *
6: /SIDSL/gcgdata/geneseq/geneseq/AA1984.DAT: *
7: /SIDSL/gcgdata/geneseq/geneseq/AA1985.DAT: *
8: /SIDSL/gcgdata/geneseq/geneseq/AA1986.DAT: *
9: /SIDSL/gcgdata/geneseq/geneseq/AA1987.DAT: *
10: /SIDSL/gcgdata/geneseq/geneseq/AA1988.DAT: *
11: /SIDSL/gcgdata/geneseq/geneseq/AA1990.DAT: *
12: /SIDSL/gcgdata/geneseq/geneseq/AA1991.DAT: *
13: /SIDSL/gcgdata/geneseq/geneseq/AA1992.DAT: *
14: /SIDSL/gcgdata/geneseq/geneseq/AA1993.DAT: *
15: /SIDSL/gcgdata/geneseq/geneseq/AA1994.DAT: *
16: /SIDSL/gcgdata/geneseq/geneseq/AA1995.DAT: *
17: /SIDSL/gcgdata/geneseq/geneseq/AA1997.DAT: *
18: /SIDSL/gcgdata/geneseq/geneseq/AA1998.DAT: *
19: /SIDSL/gcgdata/geneseq/geneseq/AA1999.DAT: *
20: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT: *
21: /SIDSL/gcgdata/geneseq/geneseq/AA2000.DAT: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	355	100.0	666	19	W62828
2	332	93.5	666	19	W62829
3	320	90.1	625	19	W62830
4	124	34.9	590	19	W62832
5	120	33.8	525	19	W62831
6	120	33.8	566	13	R20181
7	112	31.5	637	19	W62837
8	79	22.3	919	10	P93109
9	79	22.3	919	18	W14783
10	79	22.3	919	21	Y78914
11	79	22.3	1336	20	Y55933
12	77.5	21.8	919	10	P90996

13	77	21.7	154	20	Y33504
14	77	21.7	918	12	R12223
15	77	21.7	918	20	Y33491
16	74	20.8	1898	20	Y30795
17	73.5	20.7	593	19	W62835
18	72.5	20.4	1162	21	Y58500
19	71.5	20.1	669	19	W37483
20	71	20.0	2074	21	Y54319
21	70	19.7	28	19	W62841
22	69	19.4	365	18	W34971
23	69	19.4	449	19	W47176
24	69	19.4	449	21	Y98804
25	69	19.4	449	21	Y98805
26	69	19.4	449	21	Y80196
27	69	19.4	449	21	Y80197
28	68.5	19.3	1382	18	W31867
29	68	19.2	1299	21	Y58633
30	67	18.9	423	17	R91305
31	66	18.6	112	20	Y04866
32	66	18.6	126	20	Y04861
33	66	18.6	436	17	W03662
34	66	18.6	436	20	Y22342
35	66	18.6	614	16	R82630
36	65.5	18.5	2023	21	Y54320
37	65	18.3	71	20	Y09181
38	65	18.3	371	20	W73369
39	64.5	18.2	210	19	W62560
40	64.5	18.2	594	17	W00591
41	64.5	18.2	595	17	W00595
42	64.5	18.2	611	20	Y29039
43	64	18.0	156	20	Y76583
44	64	18.0	362	18	W34972
45	64	18.0	409	20	W90342

ALIGNMENTS

RESULT 1	
W62828	W62828 standard; Protein: 666 AA.
XX	
AC	W62828;
XX	
DT	27-OCT-1998 (first entry)
XX	
DE	Macadamia integrifolia antimicrobial protein.
XX	
KW	antimicrobial protein; infestation; control.
XX	
OS	Macadamia integrifolia.
XX	
FH	Key
FT	Peptide
FT	Location/Qualifiers
FT	/note= "signal peptide"
FT	29..666
FT	/note= "mature protein"
XX	
PN	W09827805-A1.
XX	
PD	02-JUL-1998.
XX	
PF	22-DEC-1997; 97WO-AU00874.
XX	
PR	20-DEC-1996; 96AU-0004275.
XX	
PA	(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX	
PI	Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
XX	
DR	WPI: 1998-377279/32.
XX	
DR	N-PSDB: V42310.
XX	

Human unliganded a
Human androgen rec
Human androgen rec
A human trichoyal
Zea mays antimicro
HHV8 ORF 73 protei
Mouse liver cancer
Amino acid sequenc
Stenocarpus sinuat
Chimeric Ewing's s
Wilms' tumour poly
Human WT1 protein
Mouse WT1 protein
Mouse Wilms' tumou
Human Wilms' tumou
Human metastasis-a
Protein regulating
Transcription fact
Mycobacterium spec
Mycobacterium spec
Human 70K UI snRNP
70K UI snRNP bindi
70K autoantigen, p
Amino acid sequenc
Peptide seq ID No:
Epitope tagged TRP
Protein of DNA mar
SAP-AlaMet-VEGF165
SAP-GlySer-VEGF165
T. gondii immunoge
Human ovarian tumo
Chimeric Ewing's s
G. max truncated S

PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT useful for controlling microbial infestations of plants or mammals
XX
XX Claim 1; Page 34-36; 96pp; English.
PS
CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian
CC animals.
XX
XX Sequence 666 AA;
SQ

Query Match 100.0%; Score 355; DB 19; Length 666;
Best Local Similarity 100.0%; Pred. No. 1.8e-32;
Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 KRDPQOREYEDCRRRCQOEPRQOHQCLRCREQOROHGRGDMNPORGSGRYEEGE 60
Db 186 krdpqreyedcrrrcceqgqprqgqclrcrcreqqrqggrgdmnpqrgsryeegee 245
OY 61 EQS 63
Db 246 eqs 248

RESULT 2

W62829
ID W62829 standard; Protein; 666 AA.
XX
XX W62829;

27-OCT-1998 (first entry)

DE Macadamia integrifolia antimicrobial protein.
XX
XX antimicrobial protein; infestation; control.
KW
XX
XX Macadamia integrifolia.
OS
XX

Key Location/Qualifiers
FH Peptide 1..28
FT /note= "signal peptide"
FT 29..666
FT Protein /note= "mature protein"

W09827805-A1.

02-JUL-1998.

22-DEC-1997; 97WO-AU00874.

20-DEC-1996; 96AU-0004275.

(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.

XX Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;

PI WPI: 1998-377279/32.

DR N-PSDB; V42311.

PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT useful for controlling microbial infestations of plants or mammals

XX Claim 1; Page 39-41; 96pp; English.

CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian
CC animals.
XX
XX Sequence 666 AA;

Query Match 93.5%; Score 332; DB 19; Length 666;

Best Local Similarity 92.1%; Pred. No. 7e-30;
Matches 58; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

OY 1 KRDPQOREYEDCRRRCQOEPRQOHQCLRCREQOROHGRGDMNPORGSGRYEEGE 60
Db 186 krdpqreyedcrrrcceqgqprqgqclrcrcreqqrqggrgdmnpqrgsryeegee 245
OY 61 EQS 63
Db 246 kqs 248

RESULT 3

W62830
ID W62830 standard; Protein; 625 AA.

W62830;

27-OCT-1998 (first entry)

DE Macadamia integrifolia antimicrobial protein.

XX antimicrobial protein; infestation; control.

OS Macadamia integrifolia.

Key Location/Qualifiers

FH Peptide 1..28

FT /note= "signal peptide"

FT Protein /note= "mature protein"

W09827805-A1.

02-JUL-1998.

22-DEC-1997; 97WO-AU00874.

20-DEC-1996; 96AU-0004275.

(RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.

XX Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;

PI WPI: 1998-377279/32.

DR N-PSDB; V42316.

PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT useful for controlling microbial infestations of plants or mammals

XX Claim 1; Page 43-45; 96pp; English.

CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian
CC animals.

XX Sequence 625 AA;

Query Match 90.1%; Score 320; DB 19; Length 625;
Best Local Similarity 88.9%; Pred. No. 1.5e-28;
Matches 56; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

OY 1 KRDPQOREYEDCRRRCQOEPRQOHQCLRCREQOROHGRGDMNPORGSGRYEEGE 60
Db 145 krdpqreyedcrrrcceqgqprqgqclrcrcreqqrqggrgdmnpqrgsryeegee 204
OY 61 EQS 63
Db 205 kqs 207

RESULT 4

```

W62832
ID W62832 standard; Protein; 590 AA.
XX
AC W62832.
XX
DT 27-OCT-1998 (first entry)
XX
DE Gossypium hirsutum antimicrobial protein.
XX
KW antimicrobial protein; infestation; control.
XX
OS Gossypium hirsutum.
XX
MN W09827805-A1.
XX
PD 02-JUL-1998.
XX
PE 22-DEC-1997; 97MO-AU00874.
XX
PR 20-DEC-1996; 96AU-0004275.
XX
PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX
PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP.
DR WPI; 1998-377279/32.
XX
PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
PT useful for controlling microbial infestations of plants or mammals
XX
PS Claim 1; Page 49-51; 96pp; English.
CC The sequence is that of an antimicrobial protein which can
CC be used to control microbial infestations in plants and mammalian
CC animals.
XX
SQ Sequence 590 AA;

Query Match 34.9%; Score 124; DB 19; Length 590;
Best Local Similarity 37.0%; Pred. No. 1.9e-06;
Matches 34; Conservative 9; Mismatches 15; Indels 34; Gaps 5.

QY 3 DPQQEVEDCCRRCBQEPPOQHOCQLRC-----REQQ-----RQHGRGDDMM--- 45
|||:||||::| ||||| ||| || :|||
DQ dpgrr-yeeqgcgcgqgeqrqpgqgrclkrfegqgsqrgfqeqqhchqeqdrpek 140
QY 46 -----NPGQSGSRYEGGEEDQS 63
|| || | || ||| ::
DB 141 kqgcvtrecrkygenpwrg--ereeeeeet 170

RESULT 5
W62831
ID W62831 standard; Protein; 525 AA.
XX
AC W62831.
XX
DT 27-OCT-1998 (first entry)
XX
DE Theobroma cacao antimicrobial protein.
XX
KW antimicrobial protein; infestation; control.
XX
OS Theobroma cacao.
XX
MN W09827805-A1.
XX
PD 02-JUL-1998.
XX
PE 22-DEC-1997; 97MO-AU00874.
XX
PR 20-DEC-1996; 96AU-0004275.

```

XX PA (RETR-) COOP RES CENT TROPICAL PLANT PATHOLOGY.
XX XX
XX PI Bower NI, Goulter KC, Green JL, Manners JM, Marcus JP;
XX XX
XX DR WPI; 1998-377279/32.
XX XX
XX PT Novel anti-microbial protein from e.g. Macadamia integrifolia -
XX PT useful for controlling microbial infestations of plants or mammals
XX PS
XX PS Claim 1; Page 47-49; 96pp; English.
CC CC The sequence is that of an antimicrobial protein which can
CC CC be used to control microbial infestations in plants and mammalian
CC CC animals.
XX XX
SQ Sequence 525 AA;

Query Match 33.8%; Score 120; DB 19; Length 525;
Best Local Similarity 40.8%; Pred. No. 4.9e-06;
Matches 29; Conservative 13; Mismatches 17; Indels 12; Gaps 5;

OY 1 KRDPQAREYEPCRRRCEDQ--EPKQHOCQLR----REQQRHGRGDDMMNPPRGSGSR 54
Db :|||: || |::|||: | |::|||: : : ||
35 erdpqr-yeyeqrcseateerqeqqecrerykxgrqee-elqrvyqcqgr 91
OY 55 YEE--GEED 62
Db :| |::||
92 cgeqqgqgrq 102

RESULT 6
R20181
ID R20181 standard; Protein; 566 AA.
AC R20181;
DE 16-APR-1992 (first entry)
DT XX
Sequence encoded by 67 kD T. cacao protein CDNA.
DE XX
Cocoa; flavour; vicilin; seed storage protein.
XX OS
Theobroma cacao.
PN WO9119801-A.
PD 26-DEC-1991.
PF 07-JUN-1991; 91WO-GB00914.
PR 11-JUN-1990; 90GB-0013016.
PA (MRSCL) MARS UK LTD.
PI Spencer ME, Hodge R, Deakin EA, Ashton S;
DR WPI: 1992-024418/03.
DR N-PDB: Q20377.

Recombinant cocoa proteins - are responsible for flavour in cocoa beans and produced in large quantities using yeast and bacterial expression vectors

Claim 4; Fig 2; 59pp; English.

The inventors claim a 67 kD and 31 kD T. cacao protein, and fragments, and encoding DNAs. The 47 kD and 31 kD proteins are derived from the 67 kD precursor. T. cacao protein CDNA was detected in a cDNA library prepared from immature cocoa beans RNA using a probe based on the AA sequence of a CNBR peptide common to the 47 kD and 31 kD polypeptides. Homology searches revealed close

XX 20-SEP-1995; 950S-0004018.
XX (MORC-) WORCESTER FOUND BIOMEDICAL RES.
PA Zamecnik PA:
XX WPI; 1997-202879/18.
DR N-PSDB; T63407.
XX
PT Oligonucleotide(s) antisense to human androgen receptor and acidic
PT REF genes - used to inhibit gene expression, for the treatment of
PT Benign prostatic hyperplasia
XX
XX Disclosure; Page 22-28; 51pp; English.
PS
CC Human androgen receptor (M14783) binds testosterone and, acting
CC at the transcriptional level, regulates the growth of normal
CC prostatic cells. Antisense oligonucleotides (see also T63200,
CC T63404-05) based on an androgen receptor cDNA clone (see also
CC T63407) can be used to prevent androgen receptor gene expression,
CC thereby inhibiting the growth or survival of prostatic cells for
CC the treatment of benign prostatic hyperplasia and prostate cancer.
XX
SQ Sequence 919 AA:

Query Match 22.3%; Score 79; DB 18; Length 919;
Best Local Similarity 32.8%; Pred. No. 0.37; Mismatches 22; Indels 2; Gaps 1;
Matches 19; Conservative 15;

OY 5 QOREYEDCRRRCQEQEPRQOHOCRLRCRQOOROHGRGDMNPNORGGSGRYEGEGEEO 62
ID Y55933 standard; Protein: 1326 AA.
DB 58 qqgqqqg--qqgqqqqqqqqqqqetpsrqqqqggedgspqahnrptlglvlddeeq 113

CC This sequence represents the human androgen receptor (AR) amino acid
CC sequence. The invention relates to a fragment of the AR corresponding to
CC amino acids 234-391 (see Y78913). The fragment mediates
CC androgen-independent activation of the AR. The androgen receptor acts as
CC a transcription factor, regulating the expression of certain
CC androgen-responsive genes. Interaction of the AR with the protein kinase
CC A signal transduction pathway involves interaction with the androgen
CC independent region. The AR fragment and peptides derived from it can be
CC used as agents for inhibiting androgen independent activation of the
CC androgen receptor, as activation domains, and as a tool for screening for
CC compounds which affect androgen-independent activation of the AR. The
CC peptides, when used in combination with androgen deprivation, effectively
CC limit androgen mediated disease progression. These diseases include
CC cancer, benign prostatic hyperplasia, hirsutism, androgenic alopecia,
CC acne, breast cancer, Kennedy disease, and especially prostate cancer. The
CC peptides and nucleic acids encoding them, are especially used for the
CC treatment of androgen-mediated diseases, especially prostate tumours in
CC patients deprived of androgen.
XX
SQ Sequence 919 AA:

Query Match 22.3%; Score 79; DB 21; Length 919;
Best Local Similarity 32.8%; Pred. No. 0.37; Mismatches 22; Indels 2; Gaps 1;
Matches 19; Conservative 15;

OY 5 QOREYEDCRRRCQEQEPRQOHOCRLRCRQOOROHGRGDMNPNORGGSGRYEGEGEEO 62
ID Y55933 standard; Protein: 1326 AA.
DB 58 qqgqqqg--qqgqqqqqqqqqqqetpsrqqqqggedgspqahnrptlglvlddeeq 113

RESULT 10
ID Y78914 standard; protein: 919 AA.
XX Y78914:
AC 23-MAY-2000 (first entry)
DT
XX Human androgen receptor (AR) amino acid sequence.
DE
XX Androgen receptor: AR; androgen-independent activation; inhibitor;
KW cancer; benign prostatic hyperplasia; hirsutism; androgenic alopecia;
KW acne; breast cancer; Kennedy disease; prostate cancer.
OS Homo sapiens.
XX
XX MO200001813-A2.
PN
XX 13-JAN-2000.
PD
XX 30-JUN-1999; 99WO-CA00604.
PF
XX 30-JUN-1998; 98US-0091871.
PR
XX (UYBR-) UNIV BRITISH COLUMBIA.
PA
PI Sadar MD, Bruchovsky N, Gout PM, Snoek R, Mawji NR:
PI WPI; 2000-182113/16.
XX
XX Novel non-androgen ligand binding peptides for inhibiting
PT androgen-independent activation of androgen receptor, used for
PT screening compounds and for treatment of androgen-mediated diseases
PT such as prostate cancer -
XX
XX Disclosure; Page 7; 32pp; English.
XX

RESULT 11
ID Y55933 standard; Protein: 1326 AA.
XX Y55933:
AC 18-FEB-2000 (first entry)
DT
XX Human ZC3 protein.
DE
XX Antirheumatic; antiarthritic; antiinflammatory; antiallergic; osteopathic;
KW antipsoriatic; antiarteriosclerotic; antiasthmatic; immunosuppressive;
KW neuroprotective; cardiact; cerebroprotective; cytosstatic; antidiabetic;
KW vulnery; SRE20; protein kinase; STLK2; STLK3; STLK4; STLK5; STLK6; STLK7;
KW ZC1, ZC2, ZC3, ZC4, KHS2, SNU1, SNU3, CFX2, PAK4; PAK5; antagonist;
KW antibody; gene therapy; rheumatoid arthritis; atherosclerosis; asthma;
KW inflammatory bowel disease; Crohn's disease; osteoarthritis; psoriasis;
KW rhinitis; autoimmunity; organ transplantation; multiple sclerosis;
KW myocardial infarction; cardiovascular disease; stroke; renal failure;
KW oxidative stress-related neurodegenerative disorder; Parkinson's disease;
KW amyotrophic lateral sclerosis; Leigh syndrome; cancer; cardiomyopathy;
KW ischemic disorder; inflammation; diabetes mellitus; fibrosis; mitosis;
KW mesangial disorder; growth regulation; wound healing; T cell activation;
KW immunosuppressant.
XX
XX Homo sapiens.
OS
XX MO9953036-A2.
PN
XX 21-OCT-1999.
PD
XX 13-APR-1999; 99WO-US08150.
PF
XX 14-APR-1998; 98US-0081784.
PR
XX (SUGE-) SUGEN INC.
PA
XX Plowman G, Martinez R, Whyte D;
PI WPI; 1999-611301/52.
XX
XX N-PSDB; Z40485.
DR
XX

CC A proapoptotic dependence domain mediated pathological conditions e.g.
CC Huntington's disease, Alzheimer's disease, Kennedy's disease,
CC Spinocerebellar ataxias, dentatorubropallidolysian atrophy,
CC Machado-Joseph disease, stroke or head trauma. They can also be used for
CC reducing the severity of a pathological condition mediated by upregulated
CC cell proliferation or cell survival e.g. neoplastic, malignant,
CC autoimmune or fibrotic conditions. This sequence represents a human
CC unliganded androgen receptor described in the method of the invention.
XX
50 Sequence 154 AA:

Query Match	21.7%;	Score 77;	DB 20;	Length 154;
Best Local Similarity	32.8%;	Pred. No. 0.1;		
Matches 19;	Conservative 12;	Mismatches 21;	Indels 6;	Gaps 1;

QY 5 QQRREYEDCRRRCEQQEPFRDQHCCCLRCREQDRQHGRGGMNPNPQRGSRGRYEEGEFEQ 62
||::: ||: || | : |||: | | ||: |
Db 58 qqqqqq-----qqqqqqqqqqlsprrqqqqggedsgspahryrptgylvlddeeq 109

RESULT	14
R12223	
ID	R12223 standard; Protein; 918 AA.
vv	

PN WO9107423-A.

Query Match 21.7%; Score 77; DB 12; Length 918;

Best Local Similarity 32.8%; Pred. No. 0.62;
Matches 19; Conservative 12; Mismatches 21; Indels 6; Gaps 1;

QY 5 QQRREYEDCCRRRCRQDQEPFRQHOCILCRREQDKHGGRGGMNPRQKSGSGRYDEGEEDQ 62
||::: ||| | : |||: | | :: |
Db 58 qqqqqq-----qqqqqqqqqsprqqqqggedsqpahrrcptgyLvldeeqq 109

RESULT	15
Y33491	
ID	Y33491 standard; Protein; 918 AA.

Query Match	21.7%	Score 77:	DB 20:	Length 918;
Best Local Similarity	32.8%	Pred. No. 0.62:		
Matches	19;	Conservative	12;	Mismatches 21;
			Indels	6;
			Gaps	1.

QY	5	QQRREDCRRRCEDQEPFRQOHCQQLRCRQQRHQRGGGDMNPPQGGSGRYEGGEDEQ	62
	58	qqqqqqq-----qqqqqqqqqqqlsprrqqqqqqqqdsspqahrrrrfpqlvyldeeqq	109

Dy 5 QRRLEDCRRRCFQEPKQHQCCLACREQQRHGRGGMMNPQRGSGRYEEGEEOO 62
 |:: : ::| | : |||| | | ::| |
Db 58 qqqqg-----qqqqqqqqqsprqqqqggedsqaahrtpgtylvIddeeq 109

Fri Mar 2 09:28:41 2001

us-09-331-631a-1_copy_186_248.rag

Page 8

Search completed: March 1, 2001, 15:47:02
Job time: 227 sec
